Diese Website verwendet Cookies von Google, um Dienste anzubieten und den Traffic zu analysieren. Deine IP-Adresse und dein User-Agent werden zusammen mit Messwerten zur Leistung und Sicherheit für Google freigegeben. So können Nutzungsstatistiken generiert, Missbrauchsfälle erkannt und behoben und die Qualität des Dienstes gewährleistet werden.

WEITERE INFORMATIONEN OK

An unending variety of topics

Thursday, May 14, 2009

Paradoxes resolved (mostly)

A paradox is usually just a good *reductio ad absurdum*, or what mathematicians call a proof by contradiction. The paradox serves to disprove one of its assumptions (or equivalently, if we "exclude the middle", to prove the opposite of one of its assumptions). Many intellectuals often overly complicate paradoxes, for the sake of entertainment or out of wishful thinking that the dubious assumption has not in fact been disproven. A paradox is only truly mysterious if more than one of the assumptions being used are truly mysterious, and thus we can't tell which assumption caused the contradiction. But usually it's more a case of obsfucation or wishful thinking than of genuine mystery. Many academics have a terrible habit of hiding or confusing their assumptions, and this allows wishful thinking to run rampant.

Newcomb's paradox

Chooser -- let's say that's you --- plays a game wherein you choose whether to open both boxes A and B or just B. For whatever odd reason -- Predictor is an omniscient diety, Predictor is running a very repeatable simulation of the Chooser's mind, time travel, the Predictor has repeatedly run the Chooser through this experiment before, etc. -- the Predictor already knows with very high probability what box(es) the Chooser will select. The first step in the game is that the Predictor makes his prediction. Then a trustworthy third party puts $1,000 into box A, and $1,000,000 into box B if the Predictor has predicted B only, otherwise nothing into box B. After these two steps are completed the Chooser selects which box(es) to open.

Here's the payoff matrix for Chooser:

Predicted choice	*Actual choice*	*Payoff*
A and B	A and B	$1,000
A and B	B only	$0
B only	A and B	$1,001,000
B only	B only	$1,000,000

What is Chooser's best strategy? According to game theory, A and B beats B-only by $1,000 whether the Predictor predicts A and B or just B. Therefore Chooser should choose to open both A and B.

Pages

- Home

About Me

Nick Szabo

"A premier thinker about history, law and economics, and the lessons they have for security." -- Adam Shostack, Emergent Chaos

"Szabo comes out with these essays that leave me in awe." -- Brian Dunbar

"Reading material that is eclectic, challenging and endlessly fascinating." -- Sean McGrath, Propylon

"Like most blogs worth my attention, this blog is updated only infrequently. That is because the authors of blogs worth my attention only post when they have something to say that is true, relevant and not already known by their audience. Most of the human race does not have the skill to know when an idea has these three properties. The skill is particularly rare in the fields of politics and economics, which is why this blog is such a rare and valuable thing." -- Richard Hollerith

View my complete profile

But another strategy for Chooser is the following: The Chooser, knowing that the Predictor has very good information about his forthcoming choice and has mechanically acted on it, and thus is almost surely correct, eliminates those choices where the Predictor is wrong and chooses B-only over A and B because of the remaining two choices it has a higher payoff.

By following 2-player game theory, the expected value of Chooser's winnings is only a bit over $1,000, while by taking into account the high accuracy of the Predictor's prediction and maximizing expected value the expected value is a bit over $1,000,000.

The paradox dissappears if you stop assuming a game with two players. The Predictor does not have any stake in the outcome and is not a player. Indeed, we are told that Predictor will make no free choice at all, but will simply mechanically predict what Chooser will do. It's the Chooser's free will in the face of a mechanical but oddly well-informed Predictor. Chooser should not be misled by the payoff matrix into assuming this is a 2-player game, but should instead choose the best of the two choices of significant probability.

Fermi's Paradox

Enrico Fermi asked, if there are extraterrestrial civilizations, why haven't we seen them? No alien artifacts large or small on earth, nor anything visibly unnatural on the many millions of square kilometers of billion-year-old (or more) surface we've observed elsewhere in our solar system, nor any visible megastructures in our galaxy. Under Darwinian evolution life and civilization tends to spread to use as much energy and matter as it can. New volcanic islands and areas where life has been destroyed by volcanoes are quickly colonized by a very observable spread of plants that soon soak up a significant fraction of the incoming sunlight. Human civilizations have similarly in the blink of astronomical time spread all over our planet, leaving a number of highly visible artifacts such as the Great Wall of China, a dazzling display of lights on the night side of our planet, and in our atmosphere increased carbon dioxide and a potpouri of odd chemicals. Photosynthetic life has given our planet a bizarre oxygen atmosphere and turned our continents significantly darker. Even if an alien society somehow turned radically un-Darwinian and thus remained obsessively small and hidden, it would take just one "crazy hermit" to appear once in hundreds of millions of years to take off and replicate his crazy Darwinian verion of that civilization across the galaxy, building visible structures all over the galaxy to efficiently use stellar energy and recycle volatile elements.

The average star in our galaxy is about 10 billion years old; if it takes 5 billion years for life to appear and evolve into a civilization and 200 million years for the their descendants to spread across our galaxy (which only requires travel at a small fraction of the speed of light), the average civilization that has already emerged in our galaxy should have spread across it 2.3 billion years ago.

If they exist, or existed, evidence of this existence should be all over the galaxy, just as with the existence of life and civilization on earth. But we see no evidence that any advanced civilization has ever expanded into our solar system or any where else in our galaxy. Even worse, we see no evidence of artificiality in other galaxies. Artificial entities would severely change what we observe in the cosmos, but instead the millions of galaxies we've observed seem to look quite natural. This can be easily tested by studying data from our current telescopes. Unless we discover a significant proportion of galaxies that are oddly dim in the optical but bright in the infrared (or perhaps, if the alien machines are extremely clever and very miserly, in the microwave), indicating artificially efficient use of stellar energy, and oddly bright in the heavy element spectra and dim in the volatile elements spectra, indicating artifically efficient enclosure of the volatile elements (for efficient volatile recycling), we must conclude that the odds of finding a civilization in any given galaxy, and thus of another civilization in our own galaxy, are remote. Fermi's paradox simply proves its major assumption wrong: there are no little green men in our galaxy. Sorry to pop that good old sense of wonder.

It would be fun to listen in on a civilization from a distant galaxy, if we ever find one and if we ever figure out how to detect such faint signals. Imagine the Wikipedia of a billion-year-old civilization!

Kavka's toxin

Perhaps less easily resolved, because one of its assumptions is the vague and subjective idea of "intent", is the paradox of Gregory

Blog Archive

- ► 2018 (1)
- ► 2017 (3)
- ► 2016 (4)
- ► 2015 (3)
- ► 2014 (3)
- ► 2013 (3)
- ► 2012 (8)
- ► 2011 (12)
- ► 2010 (9)
- ▼ 2009 (29)
  - ► November (3)
  - ► October (6)
  - ► September (3)
  - ► August (1)
  - ► July (1)
  - ► June (3)
  - ▼ May (6)
    - Sotomayor, Calabresi, and the Chief Justice at our...
    - Futarchy: an experiment we'd learn a great deal fr...
    - What caused the birth of agriculture and the indus...
    - Political property and the evolution of liberty
    - Paradoxes resolved (mostly)
    - Liar-resistant government
  - ► April (3)
  - ► January (3)
- ► 2008 (55)
- ► 2007 (47)
- ► 2006 (130)
- ► 2005 (44)

Kavka's toxin: you get $1 million put in your bank account at 9 AM if at 7 AM you intend to drink an extremely painful but not otherwise harmful toxin at 11 AM. The toxin basically causes you to live in sheer hell for 24 hours. You get to keep the million dollars whether you drink the toxin or not. Can you intend to do something that when the time comes would not be rational to do? Kavka says you can't -- that there is no way to win the $1 million. Turn off your alarm and sleep in.

My analysis is that the only ways to win the $1 million are through credible commitment or self-delusion. Thus the Wikipedia entry cites election-year political promises that would actually be too expensive to implement as an example of Kavka's toxin. Of course the political party only fools others (and perhaps themselves) into believing its intent. Standard election procedures, in which campaign promises are not legally binding, prevent credible commitment, so serious intent only could arise through self-delusion.

Posted by Nick Szabo at 12:39 AM

5 comments:

coldequation said...

Re: Kavka's toxin, I don't think self-deception would fly in general. If you can deceive yourself, that implies that you don't have a single unitary ego aware of and in control of your intentions. If that's true, then having your conscious ego intend to take the poison != "you" intending to take the poison. You can't assume a unitary ego for the purposes of getting paid, but a split personality for the purposes of carrying out the action, unless the billionaire is a complete idiot (like the electorate - but even that case is questionable because politicians may not believe that the "poison" is poisonous when they're not actually governing, or they may be flat-out lying).

Your best bet is to get somebody to really force you to do it, or to set up a "doomsday device" that will force feed it to you - a machine that feeds you the poison at the correct moment, and cannot be turned off once it's turned on.

5:44 AM

nick said...

Great comment, thanks. I observe self-delusion to be real, and indeed common, and it would be very interesting if it implied multiple egos lacking a "master ego" to control them. In any case, the game is played by an individual person, not an individual ego. A non-unitary consciousness raises a number of thorny problems. For example, can the egos control each other? If the Self-Deluder intends and the Knower knows this intent is a self-delusion, then intent might be imputed to Knower as well, depending on how the game rules or the referee define "intent".

Assuming some external evidence of intent, such as promise, it would definitely be imputed to this individual by the law, which takes no cognizance of multiple egos (at least if the person is sane and the only symptom is common self-delusion). Indeed, this is one of the vague aspects of "intent" -- is it really a known internal mental state or is it, as in law, just a statement about a class of mental states we infer from external action?. Albeit almost all of us an even some other animals are mentally wired to make such inferences: as Justice Holmes quipped, even a dog can tell the difference between being kicked and being stumbled over.

Getting somebody or something to force you to do it is basically what I meant by credible commitment. There is, however, a difference between intending to do it yourself at 11 AM and intending that it be done to you at 11 AM. You could, for example, take pills at 7 AM that release the toxin at 11 AM. But does that really constitute intending to drink the toxin at 11 AM? If it could be so construed we could more narrowly word the rules of the game, to talk only very specifically of drinking the toxin yourself from a cup at 11 AM under no coercion, and intent to do same, to eliminate most and perhaps all possibilities of credible commitment. An interesting exercise in legal drafting. :-)

11:23 AM

G Williams said...

Fermi's Paradox is actually a response to the Drake Equations, the best current solutions of which yield a universe with plentiful extraterrestrial life.

As an attempt to disprove those formulations, it has succeed reasonably well, although since Fermi first proposed the Paradox, and even since Hart expanded on it int he seventies, we have a better understanding of how invisible extraterrestrial life might be.

For starters, Population one stars, the youngest stars and the only to be metallic enough to support complex life and star-faring cultures, are relatively young, on average only about 5 billion years old (metalicity and star populations have only really be examined since the mid seventies, and our understanding of the origin of stars and planets has grown hugely since then) Using this number instead of taking the average age of all stars (many of which are completely unsuited to life) It's apparent that star-faring cultures would only be a few hundred million years ahead of us. Taking into account extinction events such as have occurred to our own planet, puts these potential civilizations on a much more even footing with our own, especially since it may take significantly longer to begin expanding than Fermi or Hart ever suspected.
So, extraterestrial life is likely to be significantly less well established than either Fermi or Hart suspected (significantly less well established than Drake suspected as well)

This new understanding of the potential age of star faring cultures means that, even without taking into account extinction events and other potential barriers to development and spreading, these civilizations are likely to be much harder to detect than Fermi and Hart suspected. In fact, only the recently launched Kepler orbital observatory is theoretically capable of detecting many of the signs of life Fermi predicted, and those only within a relatively small bubble of space. Earlier telescopes and observatories have not actually been able to detect the minimal changes in a stars emission spectrum that would be associated with even an earth sized planet, let alone whether that planet was inhabited or not, and even if they had been, until recently, there has not been a baseline to compare it to, before Kepler, only interstellar super-structure (such as Dyson Spheres and Rinworlds) building civilizations would have been detectable, whereas Kepler can now detect earth-sized planets within the habitable zone of their star and determine the chemical makeup of that planet, to detect signs of life. As Kepler completes it's mission in the next three to five years, we may very well lay the Fermi Paradox to rest (or not).

The truth is, that we are constantly re-evaluating the assumptions made by Fermi's Paradox and the Drake Equation. Kepler is just one way in which we are seeking to find these answers, and in the near future, we may find that Life is as plentiful as Drake first predicted, as rare as Fermi first predicted, or that we still do not yet understand all the complexities sufficiently to answer that question yet.

4:39 PM

Anonymous said...

Re. Kavka's toxin and the problem of meta-intention and credible threat-
G.E. Moore's paradox- 'can I believe x and simultaneously KNOW that x is untrue'- is relevant.
Moore's position was No you can't believe x. However, Loyalty and Identity politics- Nationalism, Family and other affectionate relationships- all depend on our Believing x with all our heart while knowing x is nonsense.
Introduce even a small amount of epistemological uncertainty and our idea of Knowledge soon turns into some sort of 'Justified True Belief' where True Belief has an increasingly theological feel.
Kavka and Newcomb derive part of their ability to grip because of the notion that the billionaire might be a very shrewd judge of character. He has Ashby requisite variety. He verges on omnescience. If only we could 'think like a billionaire' we too might get rich. Thus, it becomes important to fall in with his world view.
Here, our preferences change during the game BUT can we be sure we will not relapse into our comfortable old ways as the minutes tick on?
When we consider the Hegelian Struggle for Recognition- where the one more ready to risk death wins- or Abraham's intention to

sacrifice his son- where essentially the guy sees a goat and reckons that God is saying 'kill the goat already- what are you a cannibal?" we see the advantage of subjecting our intentions to an external governor.
But, perhaps, it would be enough to make that governor stochastic- i.e. hooking yourself to a poison machine with some probability of delivering the lethal dose.
This explains, in the epic age- or in tribal societies- the importance given to oracles and prophets who introduce that vital stochastic element.
Ultimately, Newcomb,Kavka, Axelrod etc- are posing dilemmas which reveal a problematic for decision theory- viz. the mind's need to see deep symmetries across agents such that (Noether's theorem) conservation laws operate. After all, Identity, too, is something conserved rather than given.
The fact that the actual environment might be dissipative means however that the such conservation is ironic.

2:38 AM

O.Yama said...

"Re: Kavka's toxin, I don't think self-deception would fly in general. If you can deceive yourself, that implies that you don't have a single unitary ego aware of and in control of your intentions."

A single unitary ego aware of and in control of your intentions implies a duality (two egos or division).

We could get the million dollars if we can end our processes of reflection and projection. Like the movie momento to some degree.

Or in other words, Eckhart tole could do it, speaking of someone that (allegedly) functions only in the present moment. This would suggest (if a person truly lived this way) they would not have the fragmented ego.

And that's really the dialogue that example brings awareness to, which is relevant to how we might construct "ai".

Anyways, I came to ask a question and this was the relevant article I could find:

What are the chances that life evolved on a planet that allows us to "leave"? It seems astronomical to me, if we think life is extremely rare. Is there something I am missing?

10:34 PM

Post a Comment

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.